



US 20050159654A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2005/0159654 A1**

Rao et al. (43) **Pub. Date: Jul. 21, 2005**

(54) **PATIENT DATA MINING FOR CLINICAL TRIALS**

Publication Classification

(76) Inventors: **R. Bharat Rao**, Berwyn, PA (US);
Satrajit Chandra Misra, Fremont, CA (US);
Linda Best, Antioch, CA (US);
Arun Krishnan, Exton, PA (US)

(51) **Int. Cl.⁷** **G06F 17/60**; A61B 5/00

(52) **U.S. Cl.** **600/300**; 705/3

Correspondence Address:
SIEMENS CORPORATION
INTELLECTUAL PROPERTY DEPARTMENT
170 WOOD AVENUE SOUTH
ISELIN, NJ 08830 (US)

(57) **ABSTRACT**

The present invention provides a system and method for selecting prospective patients for a clinical trial. In various embodiments, a clinical trials brokerage is configured to receive requests from drug companies for lists of persons meeting specified criteria for clinical trials. Patient records are retrieved from a structured computerized patient record (CPR) data warehouse populated with comprehensive patient information mined from unstructured hospital records. A list of persons for whom consent was obtained can be outputted and forwarded to the entity interested in performing the clinical trial and which requested the list. Anonymity of a patient can be maintained until the patient provides consent to participate in the clinical trial.

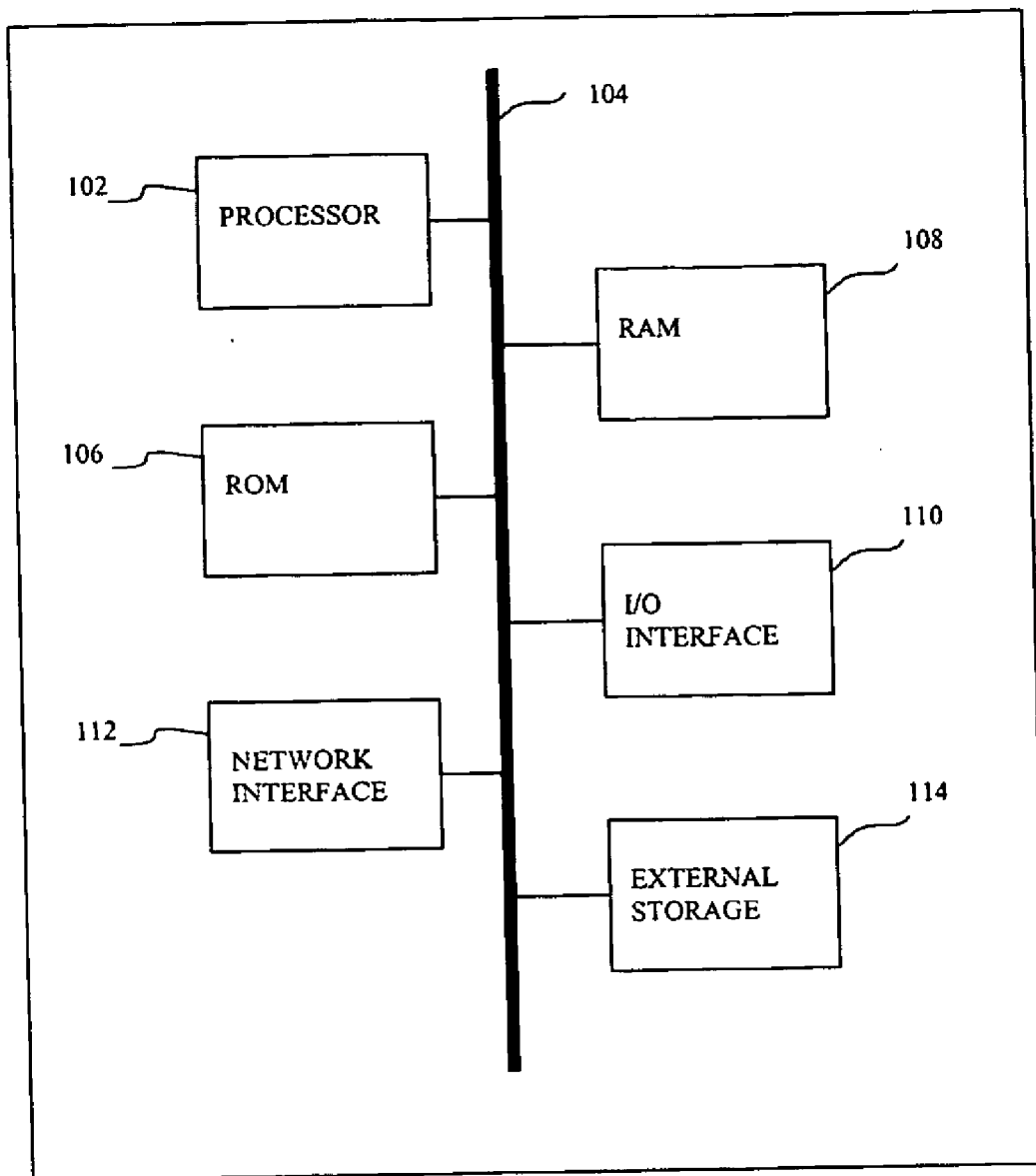
(21) Appl. No.: **11/077,728**

(22) Filed: **Mar. 11, 2005**

Related U.S. Application Data

(62) Division of application No. 10/287,098, filed on Nov. 4, 2002.

(60) Provisional application No. 60/335,542, filed on Nov. 2, 2001.



100 ↗

FIG. 1

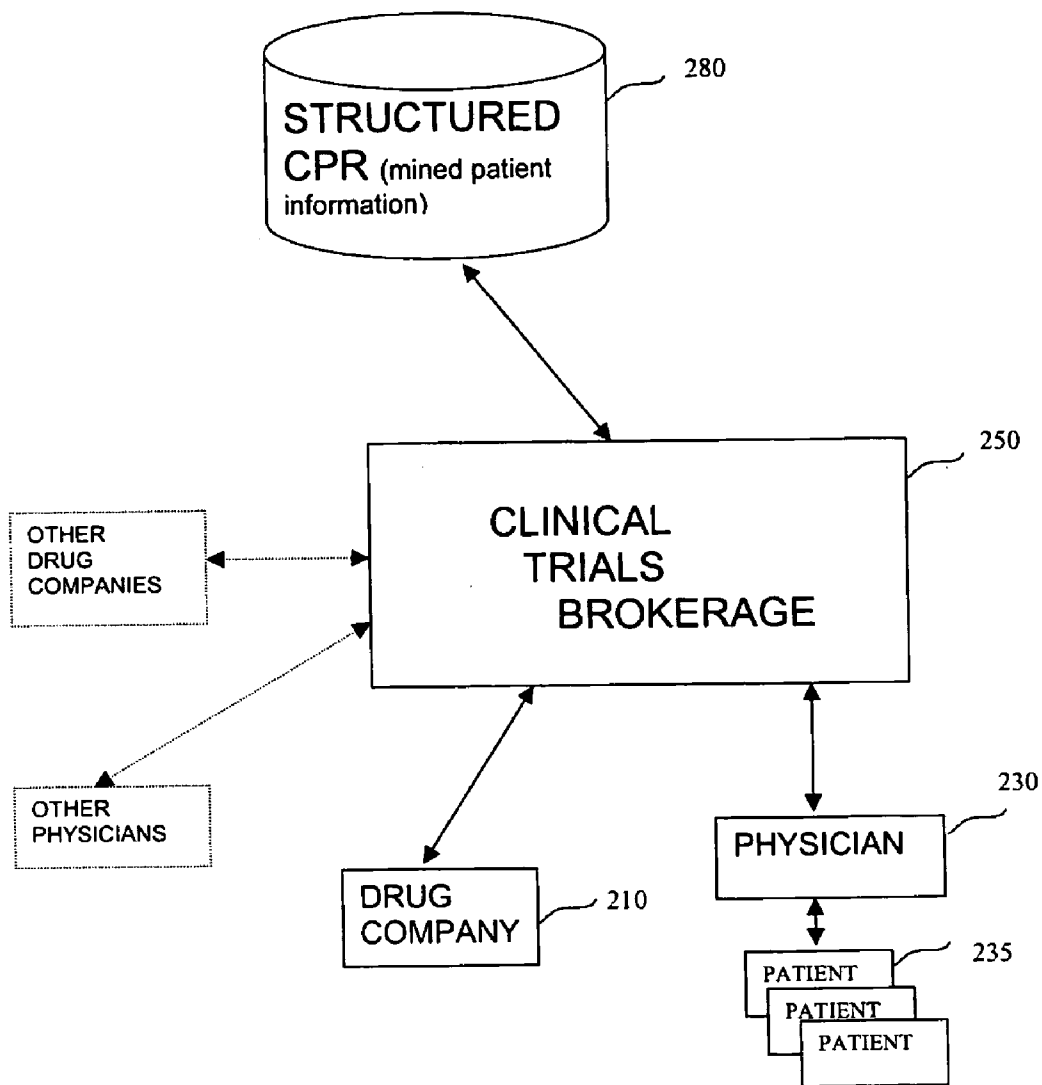


FIG. 2

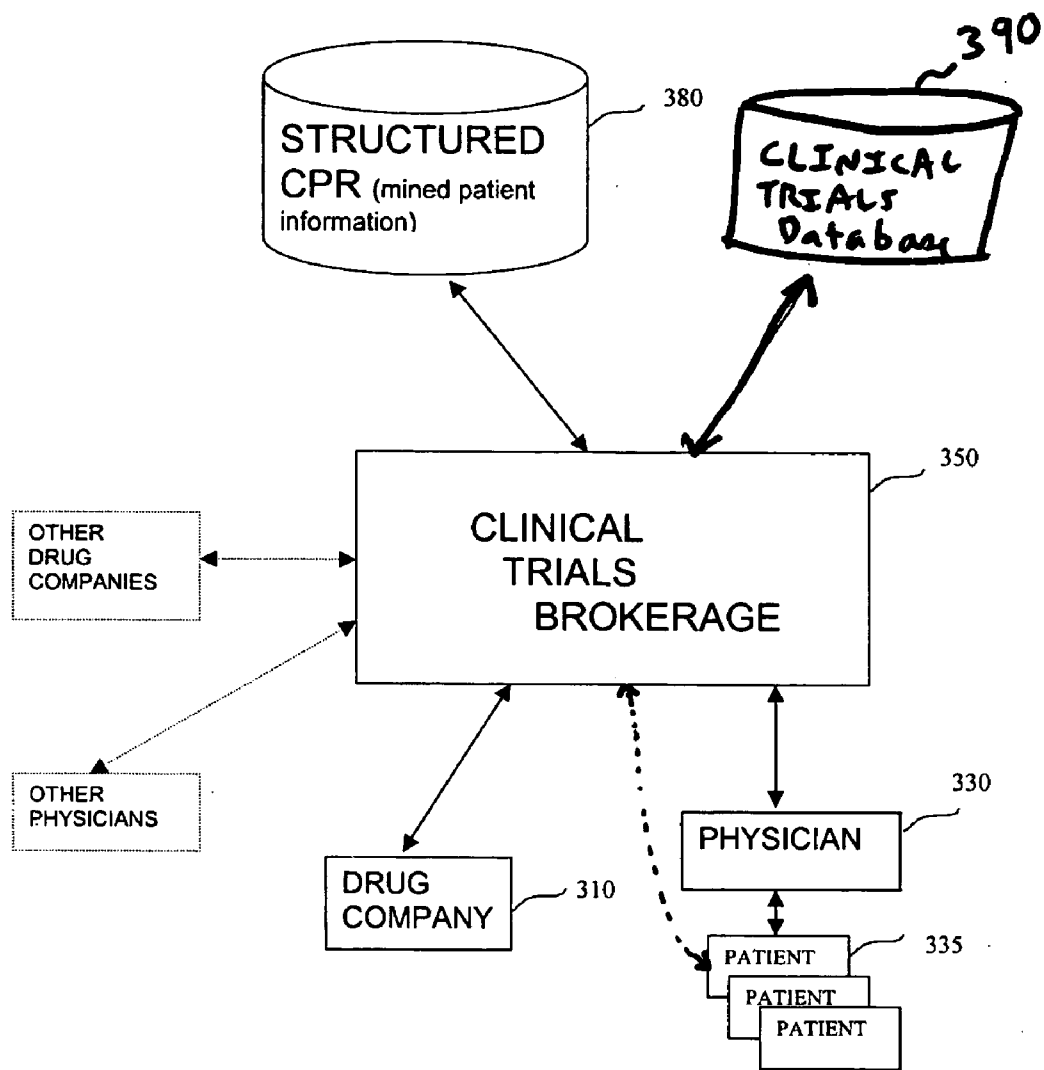


FIG. 3

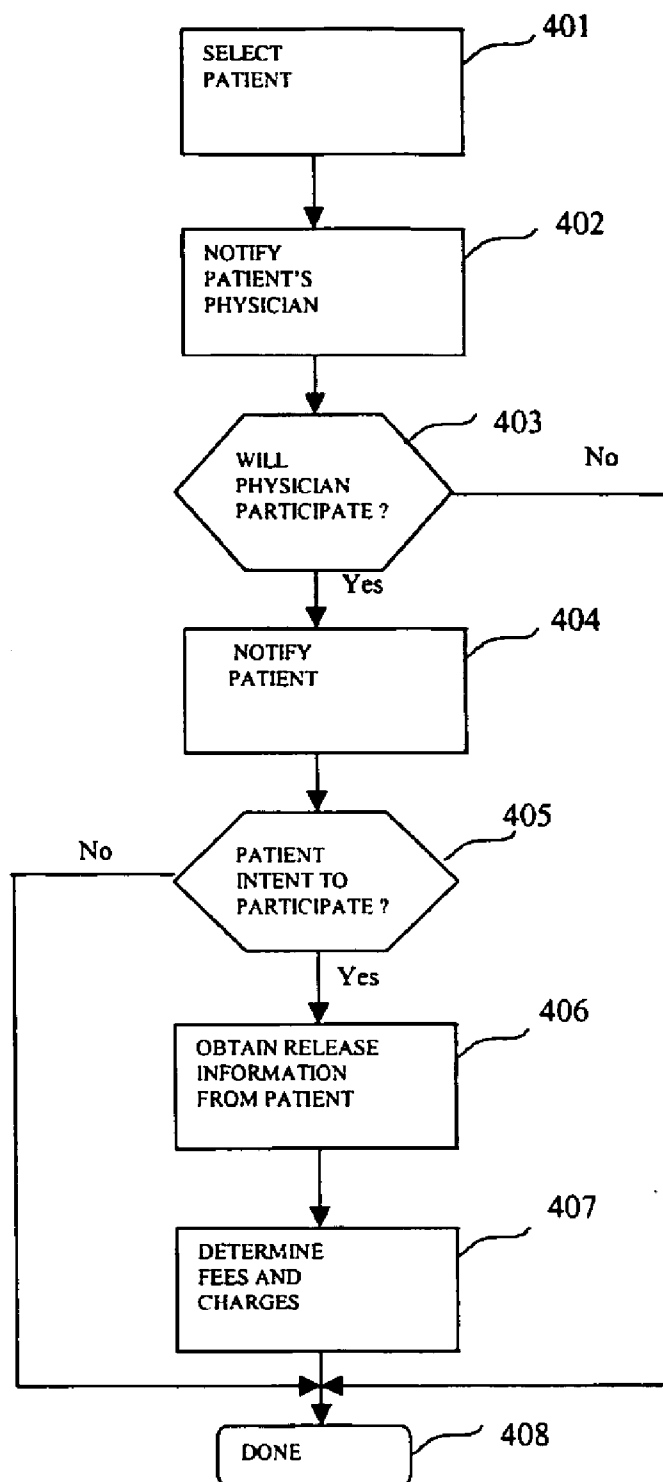


FIG. 4

PATIENT DATA MINING FOR CLINICAL TRIALS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application Ser. No. 60/335,542, filed on Nov. 2, 2001, which is incorporated by reference herein in its entirety.

FIELD OF THE INVENTION

[0002] The present invention relates to medical information processing systems, and, more particularly to a computerized system and method for selecting persons for clinical trials.

BACKGROUND OF THE INVENTION

[0003] Selection of persons for clinical trials is an expensive process. It is estimated that it costs drug companies several thousand dollars for each participant selected. Furthermore, sometimes even after being selected, persons must be dropped from a trial because of inaccurate or incorrect information. This may delay the trial, causing an even greater expense.

[0004] Although drug companies try to get the word out by placing advertisements or through direct contact with physicians, the selection process is generally quite inefficient. Physicians tend to be busy and do not always have time to respond to requests for patients, and patients may not see the advertisements for clinical trials or subscribe to the periodicals where they are placed.

[0005] Moreover, physicians at a specialized medical center tend to refer patients to trials sponsored at that center. Many physicians are unaware of all the available clinical trials because of the time it takes to keep current on all available trials for every patient that the physician sees.

[0006] In addition, clinical trials often call for very specific selection criteria and it may be difficult to ascertain if a particular person qualifies for a trial. Furthermore, because hospitals typically store information in an unstructured manner, it may be impossible using hospital records to select patients qualifying for particular clinical trials.

[0007] An equally important problem is that of matching clinical trials to specific patients. For example, for cancer alone, at any point in time there are over 600 trials in progress. Statistics show that clinical trial web sites total 75,000 hits every week, mostly from patients seeking information about trials, who are trying to get added to a trial. Estimates from National Cancer Institute indicate that only two percent of those patients eligible for a trial are in a trial. Thus, it is critically important for an individual to know if he or she may be eligible for a trial.

[0008] Given the importance and expense of selecting qualified persons for clinical trials, it would be desirable and highly advantageous to provide improved techniques for automatically selecting prospective participants for clinical trials.

SUMMARY OF THE INVENTION

[0009] The present invention provides a technique for selecting prospective participants in a clinical trial.

[0010] In various embodiments of the present invention, a method is provided that includes receiving a request for a list of prospective participants meeting specified criteria for a clinical trial. A set of patient records is then retrieved to determine persons meeting the specified criteria.

[0011] The specified criteria may include probability information, thus allowing the selection of patients likely to meet the specified criteria for the clinical study (e.g., 90% likelihood of diabetes, 70% likelihood of hypertension). In this case, the relevant patient records would include probabilistic information to allow for such selection. Additional information for each prospective participant may also be retrieved. This additional information may include information about other clinical trials that the person participated in, including whether a placebo was administered.

[0012] Furthermore, persons may still be selected even though not all information needed to determine whether a person qualifies in all respects for a clinical trial is present.

[0013] Consent to participate in a clinical trial should be obtained. A list of persons for whom consent was obtained can be outputted and forwarded to an entity interested in performing the clinical trial. Typically, this is a drug company. Physicians may be notified of their Institutional Review Board (IRB) statuses (e.g., 'approved', 'pending', or 'not approved'). Expiration dates of their status may be forwarded to approved physicians.

[0014] Because patient confidentiality is important, the anonymity of a person meeting the specified criteria must be preserved. The process of obtaining consent may include selecting physicians associated with the persons meeting the specified criteria, requesting approval to participate from each of the selected physicians, and providing consent information to persons meeting the specified criteria whose physician provided approval to participate in the clinical trial.

[0015] To further facilitate the process, questionnaires may be provided. These questionnaires may be used to ascertain qualifications for the clinical trial.

[0016] Additionally, compensation and fees can be determined for the parties involved. For example, participating physicians may be compensated. The entity requesting the list may be charged a fee. The patients participating in the clinical trial may also be compensated.

[0017] The data source used to determine the persons eligible for the clinical trial may include a data warehouse. Further, it may be populated with structured information obtained from mining unstructured patient records. The patient records may include patient information obtained from a plurality of participating health care providers, such as hospitals.

[0018] In various alternative embodiments of the present invention, a system for selecting prospective clinical trials for an individual patient is provided. The system includes a clinical trials database, a data source containing patient information, and a clinical trials brokerage for generating a list of clinical trials for patients meeting specified criteria associated with the clinical trials. At least some of the information in the data source containing patient information may be obtained from mining unstructured patient records.

[0019] These and other aspects, features and advantages of the present invention will become apparent from the following detailed description of preferred embodiments, which is to be read in connection with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0020] FIG. 1 is a block diagram of a computer processing system to which the present invention may be applied according to an embodiment of the present invention;

[0021] FIG. 2 shows an exemplary clinical trials brokerage system according to an embodiment of the present invention;

[0022] FIG. 3 shows an exemplary clinical trials brokerage system according to another embodiment of the present invention; and

[0023] FIG. 4 shows a flow diagram outlining an exemplary technique for selecting a person for a clinical trial according to an embodiment of the present invention.

DESCRIPTION OF PREFERRED EMBODIMENTS

[0024] To facilitate a clear understanding of the present invention, illustrative examples are provided herein which describe certain aspects of the invention. However, it is to be appreciated that these illustrations are not meant to limit the scope of the invention, and are provided herein to illustrate certain concepts associated with the invention.

[0025] It is also to be understood that the present invention may be implemented in various forms of hardware, software, firmware, special purpose processors, or a combination thereof. Preferably, the present invention is implemented in software as a program tangibly embodied on a program storage device. The program may be uploaded to, and executed by, a machine comprising any suitable architecture. Preferably, the machine is implemented on a computer platform having hardware such as one or more central processing units (CPU), a random access memory (RAM), and input/output (I/O) interface(s). The computer platform also includes an operating system and microinstruction code. The various processes and functions described herein may either be part of the microinstruction code or part of the program (or combination thereof) which is executed via the operating system. In addition, various other peripheral devices may be connected to the computer platform such as an additional data storage device and a printing device.

[0026] It is to be understood that, because some of the constituent system components and method steps depicted in the accompanying figures are preferably implemented in software, the actual connections between the system components (or the process steps) may differ depending upon the manner in which the present invention is programmed.

[0027] FIG. 1 is a block diagram of a computer processing system 100 to which the present invention may be applied according to an embodiment of the present invention. The system 100 includes at least one processor (hereinafter processor) 102 operatively coupled to other components via a system bus 104. A read-only memory (ROM) 106, a random access memory (RAM) 108, an I/O interface 110, a network interface 112, and external storage 114 are opera-

tively coupled to the system bus 104. Various peripheral devices such as, for example, a display device, a disk storage device (e.g., a magnetic or optical disk storage device), a keyboard, and a mouse, may be operatively coupled to the system bus 104 by the I/O interface 110 or the network interface 112.

[0028] The computer system 100 may be a standalone system or be linked to a network via the network interface 112. The network interface 112 may be a hard-wired interface. However, in various exemplary embodiments, the network interface 112 can include any device suitable to transmit information to and from another device, such as a universal asynchronous receiver/transmitter (UART), a parallel digital interface, a software interface or any combination of known or later developed software and hardware. The network interface may be linked to various types of networks, including a local area network (LAN), a wide area network (WAN), an intranet, a virtual private network (VPN), and the Internet.

[0029] The external storage 114 may be implemented using a database management system (DBMS) managed by the processor 102 and residing on a memory such as a hard disk. However, it should be appreciated that the external storage 114 may be implemented on one or more additional computer systems. For example, the external storage 114 may include a data warehouse system residing on a separate computer system.

[0030] Those skilled in the art will appreciate that other alternative computing environments may be used without departing from the spirit and scope of the present invention.

[0031] Referring to FIG. 2, a clinical trials brokerage 250 is illustrated. The clinical trials brokerage 250 is shown operatively connected to a data repository which contains patient information typically collected from one or more health care organization, such as hospitals. This data repository is called a structured clinical patient record (CPR) 280. In various embodiments of the present invention, a plurality of drug companies, such as drug company 210, request lists of persons meeting specified criteria for clinical trials. The structured CPR 280 is then consulted to obtain the lists of persons meeting the specified criteria.

[0032] The specified criteria may include probability information, thus allowing the selection of patients likely to meet the specified criteria for the clinical study (e.g., 90% likelihood of diabetes, 70% likelihood of hypertension). In this case, the relevant patient records would include probabilistic information.

[0033] Furthermore, persons may still be selected even though not all information needed to determine whether a patient qualifies in all respects for a clinical trial is present. In this case, the list would include "persons of interest" some of whom might later be excluded from participating in the clinical trial for various reasons. Information about each person meeting the selection may additionally be provided, including information about other clinical trials that the person participated in and whether a placebo was administered.

[0034] The system may keep track of a plurality of clinical trials, and maintain a list of person who were administered a placebo instead of the drug being tested. In many cases, a person is disqualified from a trial if he or she participated in

a trial for a similar drug; however, if it is determined that a placebo was administered, the system may be configured to not exclude the person. In other cases, the system would provide information about the trial(s) that the person participated in.

[0035] A physician, such as physician **230**, may be contacted if one of their patients meets the specified criteria for a clinical trial. Prior to releasing information to a drug company, it is generally necessary to obtain agreement of the patient's physician and an informed consent of the patient to participate in the trial. For example, the physician **230** may recommend to a patient that a clinical trial being conducted by the drug company **210** would be beneficial. The details of the trial may have been forwarded to the physician **230**. Furthermore, physicians may be notified of their Institutional Review Board (IRB) statuses (e.g., 'approved', 'pending', or 'not approved'). Expiration dates of their status may be forwarded to approved physicians.

[0036] The clinical trials brokerage **250** can be notified that the patient provided an intent to participate. When the necessary informed consent information is obtained, the clinical trials brokerage **250** can provide the identity of the patient (and other patient information) to the drug company **210**.

[0037] Preferably, the structured CPR **280** is populated with patient information using data mining techniques described in "Patient Data Mining," by Rao et al., Attorney Docket No. 2001P20906US01, copending U.S. patent application Ser. No. 10/287,055, filed herewith, which is incorporated by reference herein in its entirety.

[0038] That disclosure teaches a data mining framework for mining high-quality structured clinical information. The data mining framework includes a data miner that mines medical information from a computerized patient record based on domain-specific knowledge contained in a knowledge base. The data miner includes components for extracting information from the computerized patient record, combining all available evidence in a principled fashion over time, and drawing inferences from this combination process. The mined medical information is stored in a structured computerized patient record.

[0039] To determine the specified criteria for the clinical study, multiple data sources typically need to be consulted. For example, to determine whether the patient is diabetic, the system might have to examine the following information:

[0040] (a) ICD-9 billing codes for secondary diagnoses associated with diabetes;

[0041] (b) drugs administered to the patient that are associated with the treatment of diabetes (e.g., insulin);

[0042] (c) patient's lab values that are diagnostic of diabetes (e.g., two successive blood sugar readings over 250 mg/d);

[0043] (d) doctor mentions that the patient is a diabetic in the H&P (history & physical) or discharge note (free text); and

[0044] (e) patient procedures (e.g., foot exam) associated with being a diabetic.

[0045] As can be seen, there are multiple independent sources of information, observations from which can support (with varying degrees of certainty) that the patient is diabetic (or more generally has some disease/condition). Not all of them may be present, and in fact, in some cases, they may contradict each other. Probabilistic observations can be derived, with varying degrees of confidence. Then these observations (e.g., about the billing codes, the drugs, the lab tests, etc.) may be probabilistically combined to come up with a final probability of diabetes. Note that there may be information in the patient record that contradicts diabetes. For instance, the patient is has some stressful episode (e.g., an operation) and his blood sugar does not go up.

[0046] It should be appreciated that the selection of patients for clinical trials may be based on probabilistic information. Thus, a list of patients that meet the specified criteria may comprise a list of patients likely (e.g., according to a particular degree of confidence) to have met the criteria for the clinical trial.

[0047] Since it may be necessary to obtain additional information or to verify information about a participant, the clinical trials brokerage **250** may output, or otherwise provide, questionnaires. These questionnaires may be used to ascertain qualifications for the clinical trial. For example, the patient may be asked to provide a detailed family history of particular diseases.

[0048] In addition to providing a list of persons meeting the specified criteria, the clinical trials brokerage **250** may also calculate various charges and fees. For example, participating physicians may need to be compensated. The drug company may be charged a fee for the list. Additionally, participants in the clinical trial may also be compensated.

[0049] In various embodiments of the present invention, lists of persons who are pre-qualified for certain types of clinical trials may be generated. These lists of pre-qualified individuals may be made available to drug companies or other entities interested in conducting a clinical trial.

[0050] Referring to **FIG. 3**, an alternate embodiment of the present invention is illustrated. In this embodiment, a clinical trials brokerage **350** is able to access a structured CPR **380** containing mined structured patient information, and also a clinical trials database **390** containing information about various clinical trials. The information in the clinical trials database **390** may include information regarding the qualifications for clinical trials along with other information regarding the trials. A patient, such as patient **335**, may request information about a particular clinical trial. The patient may either directly access the clinical trials brokerage **350** or go through a physician, such as physician **330**. The clinical trials brokerage **330** may access the structured CPR **380** (populated with information in the same manner as the CPR **280**) to retrieve information about the patient, and attempt to match clinical trials of interest to the patient based on the medical history of the patient and available trials.

[0051] Referring to **FIG. 4**, a flow diagram outlining an exemplary technique for selecting a person for a clinical trial is illustrated. Beginning at step **401**, a person is selected from among a set of persons meeting specified criteria. This step may include receiving a request for a list of persons meeting specified criteria for a clinical trial, and retrieving a set of patient records from a data source to determine persons meeting the specified criteria.

[0052] For example, a drug company might be interested in selecting black males who are diabetic and have had a heart attack within the last three years. This might be used to test a new drug.

[0053] Using conventional approaches, satisfying the above-mentioned selection criteria could be difficult because computerized hospital databases generally do not store such information. However, by employing the data mining techniques described in "Patient Data Mining," by Rao et al., Attorney Docket No. 2001P20906US01, copending U.S. patent application Ser. No. 10/287,055, filed herewith, a structured CPR can be populated with such patient information, thus allowing this selection criteria to be satisfied.

[0054] In step 402, the person's physician can be notified that the person has been selected for the clinical trial. At this point, a hospital's Institutional Review Board (IRB) can also be notified. The physician can also be notified if IRB approval has already been granted for this trial at this site, or if he needs to wait for the IRB approval for this trial. Next, in step 403, a determination is made as to whether the physician will participate in the study. If it is determined that the physician will participate, control continues to step 404; otherwise control terminates at step 408.

[0055] In step 404, the person is notified that he or she may qualify for the clinical trial. The patient can be directly contacted, or, indirectly contacted through a physician. At this point, the patient may be given detailed information about the clinical trial. The patient may be asked for additional information, such as through a questionnaire. The questionnaire may be used to determine qualification for the study and/or as a way to obtain additional useful information.

[0056] Next, in step 405, a determination is made as to whether the person indicated a desire to participate in the clinical trial. If the person notified his or her physician of an intent to participate, control continues to step 406; otherwise control terminates at step 408.

[0057] In step 406, release information is obtained. At this point the person may be provided with a consent form or be directed to complete one provided to him by his or her physician. Any information regarding participant compensation, including reimbursements, may also be provided. Control continues to step 407.

[0058] In step 407, fees and charges may be determined. For instance, the entity requesting the list of patients may be charged an appropriate fee for the list of patients. Furthermore, the physician and trial participants may also be compensated for their participation in the study. Control continues to step 408 where the operation stops.

[0059] As shown in FIGS. 1-4, this invention is preferably implemented using a general purpose computer system. However the systems and methods of this invention can be implemented using any combination of one or more programmed general purpose computers, programmed microprocessors or microcontrollers and peripheral integrated

circuit elements, ASIC or other integrated circuits, digital signal processors, hardwired electronic or logic circuits such as discrete element circuits, programmable logic devices such as a PLD, PLA, FPGA or PAL, or the like. In general, any device capable of implementing a finite state machine that is in turn capable of implementing the flowchart shown in FIG. 4 can be used to implement this system.

[0060] Although illustrative embodiments of the present invention have been described herein with reference to the accompanying drawings, it is to be understood that the invention is not limited to those precise embodiments, and that various other changes and modifications may be affected therein by one skilled in the art without departing from the scope or spirit of the invention.

What is claimed is:

- 1. A method for maximizing patient opportunity to participate in clinical trials, the method comprising:
 - (a) maintaining by a brokerage a list of patients administered a placebo in a first clinical trial; and
 - (b) including by the brokerage the patients of the list for consideration of a second clinical trial, the second clinical trial different than the first clinical trial.
- 2. The method of claim 1 wherein (a) and (b) are performed with a machine.
- 3. The method of claim 1 further comprising:
 - (c) determining with a machine whether the patients meet criteria for the second clinical trial.
- 4. The method of claim 1 further comprising:
 - (c) retrieving patient information with a machine from multiple, different types of data sources; and
 - (d) determining whether the at least one patient meets criteria for the second clinical trial as a function of the patient information and inclusion on the list.
- 5. The method of claim 1 wherein (b) comprises including where a first drug for the first clinical trial is similar to a second drug for the second clinical trial.
- 6. A system for maximizing patient opportunity to participate in clinical trials, the system comprising:
 - a database of a list of patients administered a placebo in a first clinical trial; and
 - a machine operable to determine eligibility of the patients for a second clinical trial as a function, at least in part, of inclusion of the patients on the list, the second clinical trial different than the first clinical trial.
- 7. The system of claim 6 wherein the machine is operable to determine whether the patients meet criteria for the second clinical trial.
- 8. The system of claim 6 wherein the machine is operable to retrieve patient information from multiple, different types of data sources and is operable to determine the eligibility as a function of the patient information.

* * * * *